Full Length Article

The Ockham's razor applied to COVID-19 model fitting French data

Mirko Fiacchini [*], Mazen Alamir

Univ. Grenoble Alpes, CNRS, Grenoble INP, GIPSA-lab, 38000 Grenoble, France

ARTICLE INFO

ABSTRACT

This paper presents a data-based simple model for fitting the available data of the Covid-19 pandemic evolution in France. The time series concerning the 13 regions of mainland France have been considered for fitting and validating the model. An extremely simple, two-dimensional model with only two parameters demonstrated to be able to reproduce the time series concerning the number of daily demises caused by Covid-19, the hospitalizations, intensive care and emergency accesses, the daily number of positive tests and other indicators, for the different French regions. These results might contribute to stimulate a debate on the suitability of much more complex models for reproducing and forecasting the pandemic evolution since, although relevant from a mechanistic point of view, they could lead to nonidentifiability issues.

1. Introduction

The Covid-19 pandemic spread affecting the whole world, causing several millions of infected cases and thousands of deaths, attracted huge efforts of researchers from different scientific fields aiming at a better understanding of the virus evolution features. Besides the epidemiologists and the medical researchers, also the data scientists and the automatic control community mobilize, one of the major issue being the inference and identification of the dynamical system ruling the pandemic evolution. The virus evolution dynamics knowledge, in fact, would allow to estimate the pandemic state and predict its future evolution and thus serve as modeling tool to infer the effects of the containment measures and support the health institutions choices. A recent survey, Alamo, Reina, and Millán (2020) offers a detailed overview on the recent scientific publications regarding modeling, forecasting and controlling the evolution of Covid-19 pandemic based on data science and control theory.

Among the simplest and most popular epidemic compartmental models, recently used for fitting the data of the spread of Covid-19, there are the SIR model, first formulated in the seminal work (Kermack & McKendrick, 1927), and the SEIR model (Fang, Nie, & Penny, 2020; Kucharski et al., 2020; Wu, Leung, & Leung, 2020). The SIR model consists in simple differential equations representing the interaction between the populations of susceptible, infected, and recovered individuals, whereas the SEIR model, considers also the population of exposed individuals, for modeling the incubation period. Variations are presented in the literature to take into account other populations involved in the Covid-19 epidemic disease evolution, for instance in Crokidakis (2020), Gevertz, Greene, Tapia, and Sontag (2020) and Morato, Bastos,

Cajueiro, and Normey-Rico (2020) that included also the asymptomatic individuals, and Giordano et al. (2020a, 2020b) that introduces the SIDARTHE model, compartmental Covid-19 model composed by 8 populations and 16 parameters. Also spatial models (Brockmann & Helbing, 2013), to model the geographical interconnection between regions, and network models (El-Sayed, Scarborough, Seemann, & Galea, 2012) to represent the social interaction between individuals and populations, have been formulated to reproduce the epidemic evolution. Also data-based, model-free models have been proposed to predict the Covid-19 spread evolution (Huang, Qiao, & Tung, 2020).

Besides non-parametric approaches for model inference, Ferretti et al. (2020), Roda, Varughese, Han, and Li (2020) for instance, parametric model identification has been widely used. The problem of parametric identification for fitting the model with the available data is strongly affected by several issues, such that the time variation of the parameters (Cori, Ferguson, Fraser, & Cauchemez, 2013), the lack of full access to the data concerning the populations (McKinley, Cook, & Deardon, 2009), the presence of uncertainties affecting the model (Fang et al., 2020; Qian & Mahdi, 2020). These sources of uncertainties and lack of exact knowledge may lead to the issue of nonidentifiability (Roda et al., 2020), that consists in practice in the existence of different sets of parameters that permit to reproduce the available data. When this occurs, it is not possible to univocally identify the real set of parameters, and then to discriminate them from all the other potential solutions of the identification problem. The consequence might be a low accuracy and a poor forecasting power of the identified model. The use of simple models (Burnham, Anderson, & Huyvaert, 2011; Postnikov, 2020; Roda et al., 2020) and the validation of the identified system on

* Corresponding author.
  E-mail addresses: mirko.fiacchini@gipsa-lab.fr (M. Fiacchini), mazen.alamir@gipsa-lab.fr (M. Alamir).

several sets of data (Portet, 2020), from different regions for instance, are possible solutions to the nonidentifiability issue.

This paper is guided by the Ockham's razor, or principle of parsimony, criterion expressible as *"Numquam ponenda est pluralitas sine necessitate"* (i.e. plurality is never to be posited without necessity). The principle of parsimony in the scientific context substantially claims that the simplest explanations of a phenomenon should be preferred to the more complex ones, since usually based on less hypotheses and then less falsifiable. As stated above for the nonidentifiability problem, indeed, the simpler is the model able to fit and reproduce the available data, the higher is expected to be its reliability in terms of accuracy and forecasting power. Starting from the available data concerning the Covid-19 pandemic evolution in the 13 regions of mainland France, an extremely simple dynamical model has been formulated with only two states and two parameters, common to all the regions. The model has been identified on the measures of deaths caused by the Covid-19 and then validated on several other time series, concerning hospitalizations, intensive care accesses, emergency admissions etc. for all the 13 regions.

Two conclusions can be drawn by our study. The first concerns the fact that the available data, although composed by several measures of the pandemic evolution in different regions, seem not to be rich enough to identify complex models. Indeed, if a two-dimensional model with two parameters permits to satisfactorily reproduce the available data, then any additional dynamics might lead to identifiability problems, since their evolution are not observable from the data. This means that different sets of parameters might generate the same measured evolution, leading to *de facto* nonidentifiability and then poor prediction capability of the model.

Moreover, the identified model results to strongly rely on the precise identification of a state related to the reproduction number evolution during the specific period of lockdown. Thus, a second, and non-surprising, conclusion is that the knowledge in advance of this evolution is necessary to predict the pandemic future behavior. On the other hand, since it seems unlikely that the data recollected during a specific period (the lockdown for instance) can exactly determine and characterize the effects of future political choices and social behavior adaptations, then any model identification based only on the data recollected during the lockdown might be unavoidably affected by poor forecasting power and it indicates the need of further knowledge.

The paper is organized as follows: Section 2 presents the time series concerning the excess of deaths due to Covid-19, used to identify the model; Section 3 deals with the model and its fitting to deaths excess data; Section 4 regards the validation of the identified model on other times series. The paper terminates with some conclusions.

2. Deaths excess data for fitting

Several datasets concerning the evolution of the Covid-19 in France are available that can be used to identify the dynamics parameters. The data are organized by gender, by department, by age and by region. This work focuses on the dynamics of the Covid-19 evolution within the different regions of mainland France, aiming at identifying a single model, as simple as possible, that permits to fit and reproduce the regional evolution of Covid-19 pandemic.

Concerning the number of demises due to the Covid-19, in particular, three time series are available. Firstly, the INSEE, the French National Institute of Statistics and Economic Studies, provides the data of all the daily demises occurred since January 1, 2018, in a single file, organized by region, in (insee.fr, https://www.insee.fr/fr/statistiques/4487988?sommaire=4487854). In Fig. 1 the evolution of deaths in Auvergne-Rhône-Alpes region, concerning the years 2018, 2019 and 2020 are represented. Besides the time series, the smoothed series, computed by averaging over a 14-days-long window, are drawn in bold. The data concerning the 13 regions of mainland France can be found in Fiacchini and Alamir (2020).

A measure of the daily deaths for Covid-19 is the difference between the daily demises occurred in 2020 and the average of those of 2018 and 2019. This difference, in fact, is supposed to represent (via some correction that are mentioned in the sequel) the excess of deaths occurred in the period of Covid-19 pandemic spread in France and can be supposed to be the effect of the epidemic.

Other two time series are available that represent measures of the demises caused by the Covid-19, in particular the number of deaths registered at the hospitals, incid_dc and the number of demises that have been electronically registered and transmitted to the Inserm[1] institute. More details on these two time series can be found in Section 4. For obtaining a reasonably reliable guess of the number of deaths caused by Covid-19, these two signals have been smoothed by averaging them over a 14-days-long time window.

The time series considered for model fitting, depicted in Fig. 2 for Auvergne-Rhône-Alpes region (see Fiacchini & Alamir, 2020 for the other regions), are:

- the excess of deaths of 2020 with respect to the average of 2018 and 2019, averaged over a 14-days-long window, denoted mean excess20;
- the average over a 14-days-long window of hospital demises time series incid_dc, denoted mean incid_dc;
- the average over a 14-days-long window of demises registered by Inserm inserm_dc, denoted mean Inserm.

From the inspection of Fig. 2, and considering the excess of deaths of 2020 as a reliable measure of the real amount of demises caused by Covid-19, it can be inferred that:

- the cases registered in the hospitals and those certified to Inserm represent a sub-estimation of the real number of demises due to Covid-19 in the first weeks after the lockdown of March 17;
- the pandemic in Auvergne-Rhône-Alpes might have been starting at the beginning of March, as the curve, although at negative values (i.e. deaths in 2020 below the average of 2018/2019), presents an exponentially increasing behavior since the first days of March;
- the lockdown might have affected also the other causes of mortality (for instance due to the traffic reduction, working activity partial stop, etc.) leading to values of excess around 0 in the last week of April and in May, when the certified demises, both in hospitals and through electronic certification, are higher than the mortality excess.

The last consideration, in particular, seems to indicate that the mortality excess of 2020 represents an underestimation of cases of demises due to Covid-19, particularly in the last part of the period taken into account. To compensate this effect, that can be observed in the data concerning many of the mainland France regions, see Fiacchini and Alamir (2020), an heuristic criterion has been employed, to avoid that the time series considered as a reliable measure of the deaths caused by Covid-19 is lower than the number of observed cases.

In particular, a linear function of the time has been added to the time series mean excess20 and the maximum between it and the signal mean incid_dc is taken. The resulting time series mean excess20 corr in Fig. 3 has been employed to identify the system.

3. Model fitting the deaths excess

A first hypothesis is formulated to express the fact that the demises caused by Covid-19 is proportional to the number of infected individuals.

[1] French National Institute of Health and Medical Research.
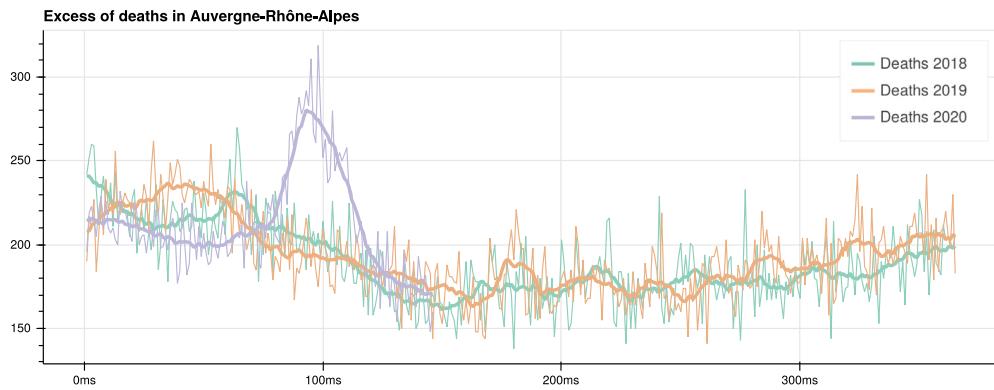
Fig. 1. Evolution of daily demises in 2018, 2019 and 2020, in Auvergne-Rhône-Alpes region. In bold, the average over a 14-days-long window.
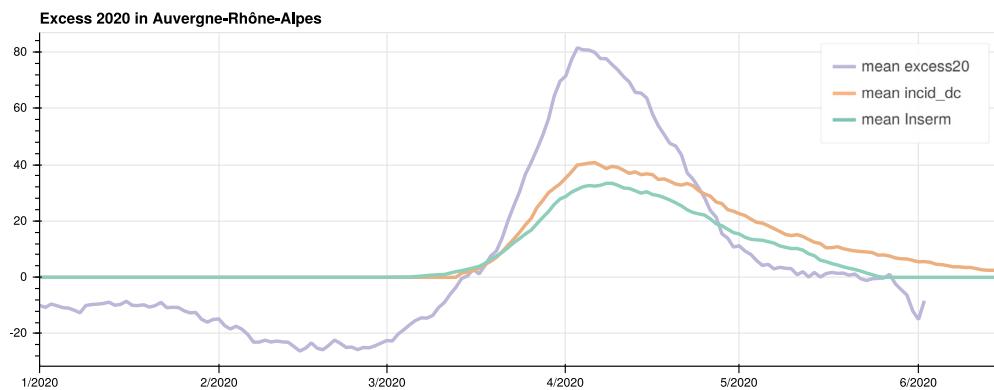

Fig. 2. Excess of death in 2020 with respect to the average of 2018/2019, `mean excess20`; deaths certified in hospitals `mean incid_dc` and deaths electronically certified to Inserm `mean Inserm`, averaged over a 14-days-long window, in Auvergne-Rhône-Alpes region.
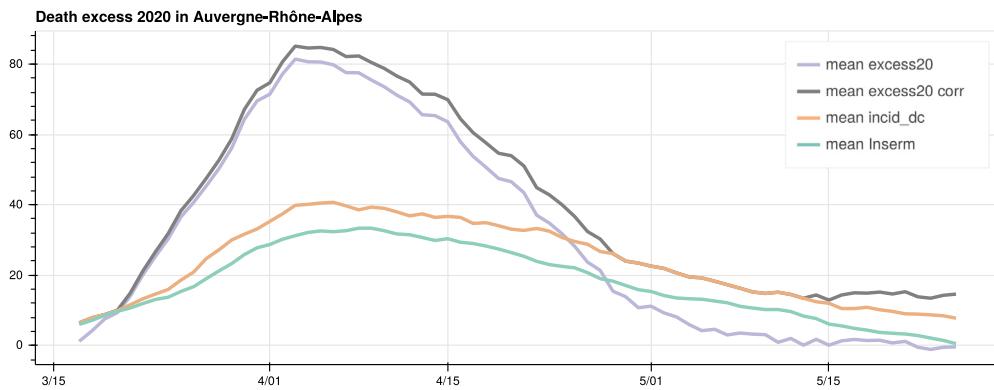

Fig. 3. Excess of death in 2020 with respect to the average of 2018/2019, `mean excess20`; deaths certified in hospitals `mean incid_dc` and deaths electronically certified to Inserm `mean Inserm`, averaged over a 14-days-long window. Corrected average of excess `mean excess20 corr`, in Auvergne-Rhône-Alpes region.

Assumption 1. The number of daily deaths due to Covid-19 is proportional to the number of infected individuals.

Consider the dynamics relating the number of Covid-19 deaths and the number of infected cases employed in several epidemiological compartmental models:

$$\frac{dD(t)}{dt} = \alpha I(t). \tag{1}$$

Assumption 1 is equivalent to suppose satisfaction of (1), often employed in the compartmental epidemic models, and then that the number of new demises for Covid-19 is proportional to the value of the infected cases, with proportionality coefficient α. Thus, a reliable

measure of the daily demises would provide a scaled image of the Covid-19 infected cases.

Moreover, as a second hypothesis, the resulting signal `mean excess20 corr` in Fig. 2 is considered hereafter as a reasonably reliable guess of the amount of daily demises due to Covid-19 infection.

Assumption 2. The time series `mean excess20 corr`, denoted $f(t)$, is a reliable measure of the number of daily demises caused by Covid-19.

Under these assumptions, then, $f(t)$ is a measured approximation of the value of the function $\alpha I(t)$, from (1), in fact:

$$\frac{dD(t)}{dt} = \alpha I(t) \simeq f(t). \tag{2}$$

Given the time series $f(t)$ that represents the evolution of the population of Covid-19 infected individuals, multiplied by an unknown constant α, its dynamics can be identified.

The simplest epidemic compartmental models, yet considered among the most robust from the information theory point of view, see Burnham et al. (2011) and Alamo et al. (2020), is the SIR model. The SIR model, and its variations SEIR and SIRD, have been recently used for fitting the spread of Covid-19 (Fang et al., 2020; Kucharski et al., 2020; Wu et al., 2020). In SIRD models, the populations of susceptible (S), infected (I), recovered (R) and dead (D) individuals are modeled by the differential equations:

$$\frac{dS}{dt} = -\beta I S$$
$$\frac{dI}{dt} = \beta I S - \gamma I - \alpha I$$
$$\frac{dR}{dt} = \gamma I$$
$$\frac{dD}{dt} = \alpha I \tag{3}$$

To pursue the aim of fitting the data with simple model, though, consider the evolution of the state variable defined as $f(t)$ in (2). From $f(t) = \alpha I(t)$ and the dynamics of $I(t)$ in (3), then

$$\frac{df(t)}{dt} = \beta(t)f(t)S(t) - \gamma f(t) - \alpha f(t) = \delta(t)f(t) \tag{4}$$

with $\delta(t) = \beta(t)S(t) - \gamma - \alpha$. Note that $\beta(t)$, that is the coefficient of the rate of infection, has been considered time-varying. In fact, the variation of susceptible individuals, i.e. the whole region population (more than 8 million people in Auvergne-Rhône-Alpes) at the beginning of the pandemic, might not be the cause of the excess of deaths drastic decreasing. This demises drop should most likely be caused by the containment measures, the lockdown in particular, that reduced the rate of infection.

Concerning, the evolution of $\delta(t)$, related to the reproduction number, a first analysis of the data showed a behavior typical of first order non-autonomous linear systems, evolving exponentially from positive values at the beginning of the pandemic expansion to negative ones during its decreasing. Thus, denoting with

- F the set of indices of the 13 regions of mainland France[2];
- $f_i(t)$ the time series of the corrected excess of deaths in 2020, obtained as illustrated above, for the ith region with $i \in F$;
- $\delta_i(t)$ the parameter of Eq. (4), for the ith region with $i \in F$;

we shall prove that the following simple two-parameter model can quite nicely account for the measurements concerning the 13 regions in the mainland France:

$$\frac{df_i(t)}{dt} = \delta_i(t)f_i(t)$$
$$\frac{d\delta_i(t)}{dt} = a\delta_i(t) + u \tag{5}$$

for every region i, with two common parameters, a and u. The objective of the identification is to find the optimal values of a and u, common to the dynamical systems of all the 13 regions of mainland France, and the initial conditions $f_i(t_0)$ and $\delta_i(t_0)$, for every ith region, such that the simulation of (5) fits the signal `mean excess20 corr`.

Hence, denoting:

- k_0 the initial day of the identification interval, i.e. the lockdown date of March 17th;
- k_f the final day of the identification interval, April 28th;

the following problem is posed:

$$\min_{a,u,\hat{f}_{i,0},\delta_{i,0}} \sum_{i \in F} \sum_{k=k_0}^{k_f} q_i(\hat{f}_i(k) - f_i(k))^2$$

$$\text{s.t.} \quad \hat{f}_i(k+1) = (1 + \delta_i(k))\hat{f}_i(k), \qquad \forall k_0 \le k \le k_f, \; \forall i \in F,$$
$$\delta_i(k+1) = (1 + a)\delta_i(k) + u, \qquad \forall k_0 \le k \le k_f, \; \forall i \in F,$$
$$\hat{f}_i(k_0) = \hat{f}_{i,0}, \qquad \forall i \in F,$$
$$\delta_i(k_0) = \delta_{i,0}, \qquad \forall i \in F. \tag{6}$$

Problem (6) is a least squares minimization problem with nonconvex constraints and 28 variables, that are:

- the two parameters a and u;
- the 26 initial conditions $\hat{f}_{i,0}$ and $\delta_{i,0}$, with $i \in F$, for the 13 regions.

Parameters q_i are chosen to appropriately weight the errors of regions with different values of excess of deaths in 2020.

To solve the nonconvex problem (6), first an initial suboptimal solution has been obtained by solving a convex problem involving only the dynamics of δ_i, with variables a, u and $\delta_{i,0}$. This solution, together with an *ad-hoc* selection of $\hat{f}_{i,0}$, has been used as initial condition of the nonconvex optimization problem. In particular, the Torczon algorithm has been employed, that is a derivative free nonconvex optimization algorithm and whose python implementation can be found at Alamir.

The optimal values for the parameters a and u in (5) obtained are

$$a = -7.1139 \cdot 10^{-2}, \qquad u = -6.2489 \cdot 10^{-3}, \tag{7}$$

and the optimal initial conditions are given in Table 1. Notice that the initial conditions of f_i for the two regions most affected by the Covid-19 spread, i.e. Île-de-France (11) and Grand Est (44), are around 40 units, much higher than those of the other regions.

The results of simulating system (5) with parameters (7) and initial conditions as in Table 1, denoted `estimation ID NL`, are illustrated in Figs. 4–10. Notice that the y axis in all the figures ranges between 0 and 200, except Île-de-France that reaches 450. The fitting, obtained with a two-parameters two dimensional system, is remarkable.

3.1. Discussion on identifiability and forecast

Although the model that proved to reproduce the evolution of the Covid-19 spread in the 13 French regions is extremely simple, with only two parameters and two states, the system involving the infected cases is nonidentifiable. In fact, it is easy to see that every pair of time series $I_i(t)$ and parameter α such that

$$f_i(t) = \alpha I_i(t)$$

for the $i \in F$ region, leads to the same trajectory of system (5). Then, from the observation of the 2020 deaths access only it is not possible to identify the real value of α and, at the same time, the time series of the infected population.

The perspective is even worse for what concerns the forecast power of system (5). Indeed, the state evolution depends on the value of $\delta_i(t)$, that can be seen as a time-varying parameter. In the case under

Table 1
Optimal initial conditions for (5) with region $i \in F$.

$\delta_{84,0}$	$\delta_{27,0}$	$\delta_{24,0}$	$\delta_{44,0}$	$\delta_{32,0}$	$\delta_{11,0}$	$\delta_{75,0}$
0.3221	0.2667	0.3598	0.2054	0.3032	0.3211	0.3167
$f_{84,0}$	$f_{27,0}$	$f_{24,0}$	$f_{44,0}$	$f_{32,0}$	$f_{11,0}$	$f_{75,0}$
7.0714	6.277	1.7143	42.8393	7.9161	37.0938	1

$\delta_{76,0}$	$\delta_{52,0}$	$\delta_{53,0}$	$\delta_{94,0}$	$\delta_{28,0}$	$\delta_{93,0}$
0.295	0.3088	0.2284	0.1204	0.3737	0.2923
$f_{76,0}$	$f_{52,0}$	$f_{53,0}$	$f_{94,0}$	$f_{28,0}$	$f_{93,0}$
3.4643	2.3189	1.8857	1.3255	1.2857	5.2883

---

[2] Île-de-France (11), Grand Est (44), Auvergne-Rhône-Alpes (84), Hauts-de-France (32), Provence-Alpes-Côte d'Azur (93), Bourgogne-Franche-Comté (27), Occitanie (76), Pays de la Loire (52), Centre-Val de Loire (24), Bretagne (53), Normandie (28), Nouvelle-Aquitaine (75) and Corse (94).

Fig. 4. Comparison of time series `mean excess20 corr` and trajectory of (5) with optimal parameters, denoted `estimation ID NL`, for Île-de-France (11) and Grand Est (44) regions.
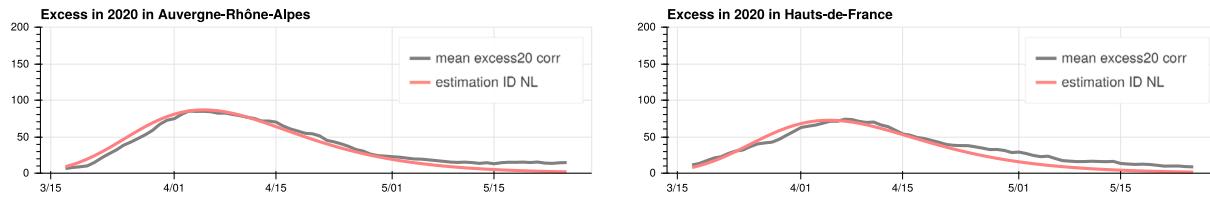

Fig. 5. Comparison of time series `mean excess20 corr` and trajectory of (5) with optimal parameters, denoted `estimation ID NL`, for Auvergne-Rhône-Alpes (84) and Hauts-de-France (32) regions.
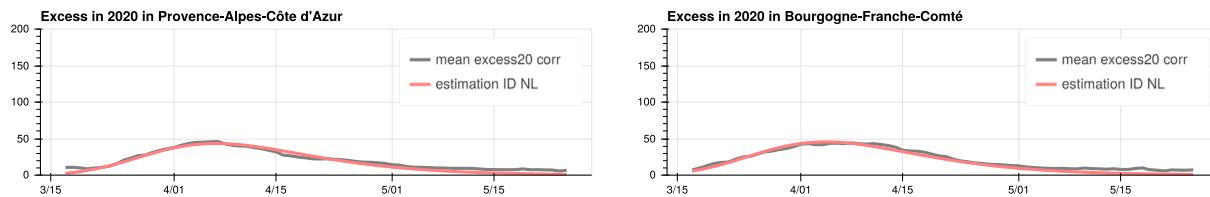

Fig. 6. Comparison of time series `mean excess20 corr` and trajectory of (5) with optimal parameters, denoted `estimation ID NL`, for Provence-Alpes-Côte d'Azur (93) and Bourgogne-Franche-Comté (27) regions.
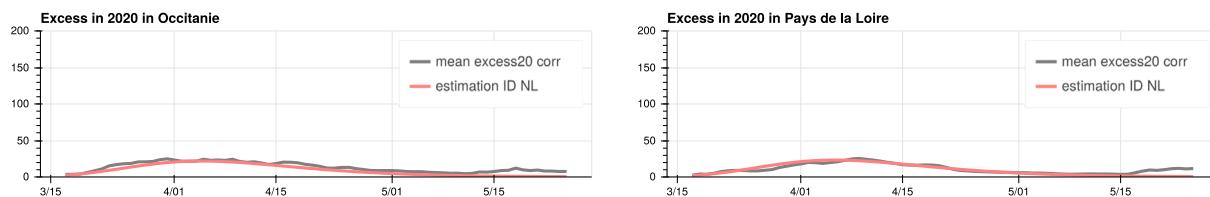
Fig. 7. Comparison of time series `mean excess20 corr` and trajectory of (5) with optimal parameters, denoted `estimation ID NL`, for Occitanie (76) and Pays de la Loire (52) regions.
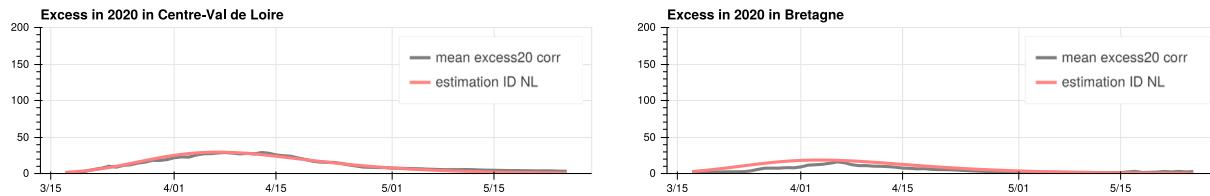
Fig. 8. Comparison of time series `mean excess20 corr` and trajectory of (5) with optimal parameters, denoted `estimation ID NL`, for Centre-Val de Loire (24) and Bretagne (53) regions.
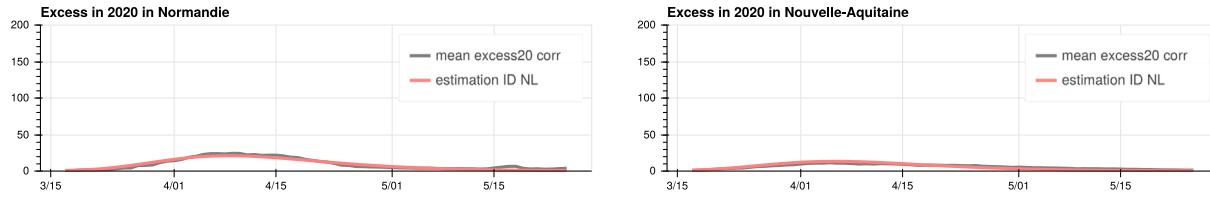

Fig. 9. Comparison of time series `mean excess20 corr` and trajectory of (5) with optimal parameters, denoted `estimation ID NL`, for Normandie (28) and Nouvelle-Aquitaine (75) regions.
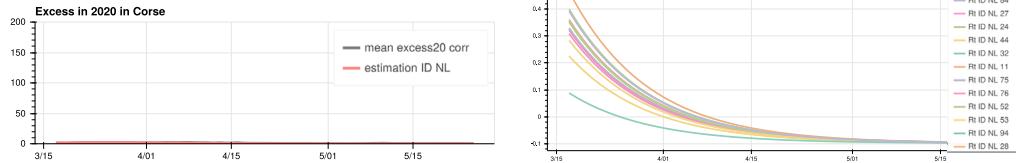
Fig. 10. Comparison of time series `mean excess20 corr` and trajectory of (5) with optimal parameters, denoted `estimation ID NL`, for Corse (94) region, and evolutions of $\delta_i(t)$.
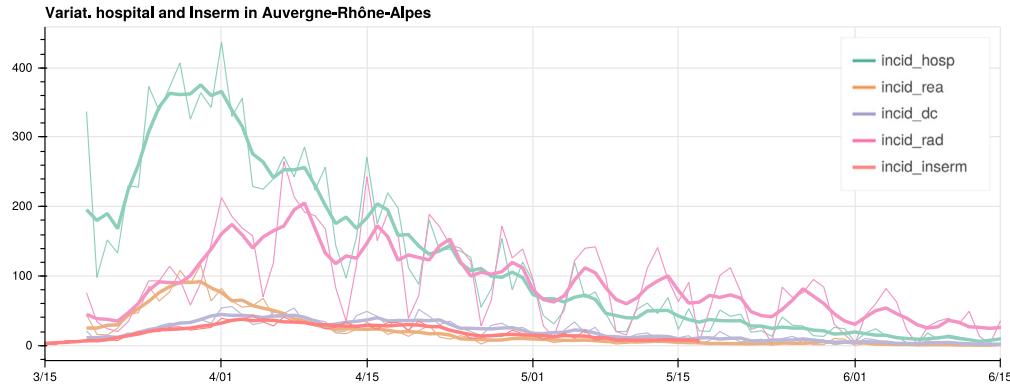

Fig. 11. Daily evolution of: new hospitalizations (`incid_hosp`); patients entering in intensive care (`incid_rea`); patient demises (`incid_dc`); recovered patients (`incid_rad`) and electronically certified demises, (`incid_inserm`), in Auvergne-Rhône-Alpes. In bold, the average over a 5-days-long window.

analysis, its evolution, caused most likely by the containment measures, the lockdown application and the people habits adaptation, has been identified with satisfactory precision. This would imply, then, that the effect of complete lockdown seems to be modeled by the specific values of a and u, resulting from the optimization process.

On the other hand, the model is inherently unable to forecast the pandemic spread unless the evolution of $\delta_i(t)$ could be predicted, since the system modeling the evolution of deaths:

$$\frac{d f_i(t)}{dt} = \delta_i(t) f_i(t)$$

leads to exponentially growing error in the prediction of the state $f_i(t)$ for $\delta_i(t)$ positive, i.e. during the pandemic spread, unless $\delta_i(t)$ is perfectly known in advance.

Thus, either one is able to predict the effect on the evolution of $\delta_i(t)$ of different containment measures and of the adaptation in the people habits, or the forecast might quickly lead to very high prediction error. On the other hand, since it seems unlikely that the data recollected during a specific period (lockdown for instance) can exactly determine and characterize the effects of future choices, we can conclude that any model identification based on the data recollected during the lockdown might be unavoidably affected by poor forecasting power. Further knowledge, on historic data, epidemiology and clinical data, as done for influenza pandemic for instance (Nishiura, Klinkenberg, Roberts, & Heesterbeek, 2009), might help to achieve this crucial task.

4. Model validation on further data

In the previous sections, the model has been fitted with the data regarding the daily deaths that can be considered as consequences of the Covid-19 pandemic. As several other datasets are available from governmental sources, though, that might be related to the pandemic evolution, we aim at check whether simple relations between these data and the identified model exist. In the affirmative case, this would validate the fitted model.

As said above, many time series regarding the Covid-19 daily evolution in France are available. In particular, the daily data on the Covid-19 can be found in (data.gouv.fr, https://www.data.gouv.fr/

fr/datasets/donnees-hospitalieres-relatives-a-lepidemie-de-covid-19/), concerning the period between mid-March and mid-June 2020, regarding:

- the number of new hospitalizations, signal `incid_hosp`;
- the number of patients entering the intensive care services, denoted `incid_rea`;
- the number of patient demises, `incid_dc`;
- and the number of recovered patients, `incid_rad`.

Fig. 11 shows the evolution of the time series in the region Auvergne-Rhône-Alpes, in thin lines the data from the files are depicted, while in bold the signals smoothed by averaging the values over 5 days are drawn.

Moreover, Fig. 11 shows also the time series, (data.gouv.fr, https://www.data.gouv.fr/fr/datasets/donnees-de-certification-electronique-des-deces-associes-au-covid-19-cepidc/#_), of the deaths for Covid-19 that have been electronically certified by the CepiDc laboratory, which is the Inserm laboratory in charge of elaborating the statistics regarding the medical demises. This time series, denoted `incid_inserm` in Fig. 11, concerns the deaths occurred in public or private hospitals, nursing homes, or other places, that have been electronically registered. This time series considers also part of the cases included in `incid_dc`. Recall that the average of time series `incid_dc` and `incid_inserm` have been employed in the definition of the signal `mean excess20 corr`, on which the model has been fitted, also see Figs. 2 and 3.

The site of the French government gives also the access to the data concerning the daily Covid-19 tests realized in every region and to the number of positive tests, see (data.gouv.fr, https://www.data.gouv.fr/fr/datasets/donnees-relatives-aux-tests-de-depistage-de-covid-19-realises-en-laboratoire-de-ville/). The evolution of the time series concerning the Auvergne-Rhône-Alpes region, smoothed by averaging it over a 5-days-long window, are depicted in Fig. 12.

At the site (data.gouv.fr, https://www.data.gouv.fr/fr/datasets/donnees-des-urgences-hospitalieres-et-de-sos-medecins-relatives-a-lepidemie-de-covid-19/), files are available reporting, for every French region and every day since mid-February, the data concerning the accesses to the hospital emergency services and the activity of SOS Médecins,
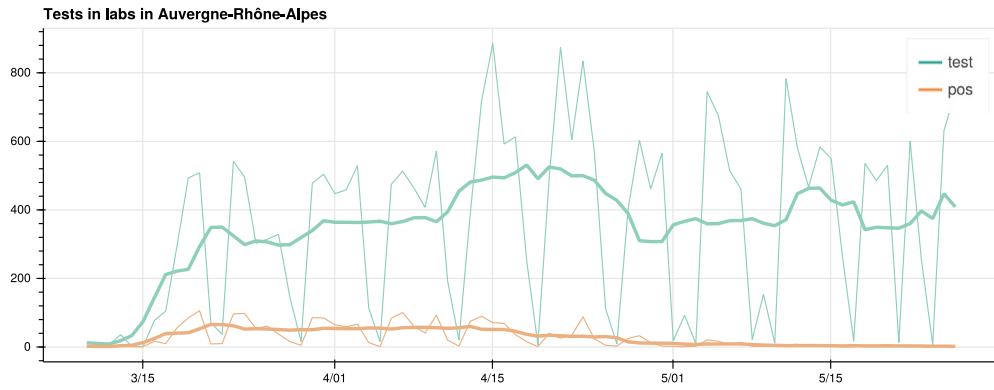
Fig. 12. Daily evolution of Covid-19 test (`test`) and tests resulted positive (`pos`), in Auvergne-Rhône-Alpes. In bold, the average over a 5-days-long window.
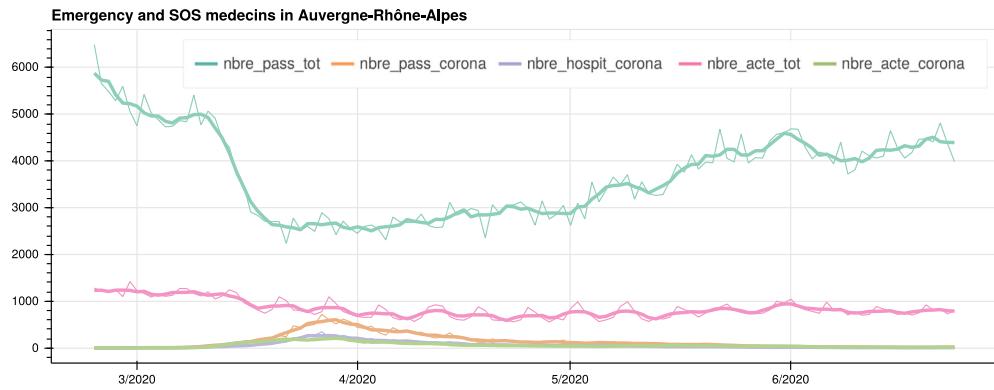

Fig. 13. Daily evolution of: admissions to emergency, `nbre_pass_tot`; admissions to emergency for Covid-19, `nbre_pass_corona`; hospitalizations among the admissions to emergency for Covid-19, `nbre_hospit_corona`; medical acts of SOS Médecins, `nbre_act_tot`; medical acts of SOS Médecins of Covid-19 cases, `nbre_act_corona`. In bold, the average over a 5-days-long window.

an association of doctors providing emergency assistance. In particular, the data series regard:

- the overall number of admissions to hospital emergency services, `nbre_pass_tot`;
- the number of admissions to emergency for presumed Covid-19 cases, `nbre_pass_corona`;
- the number of hospitalizations among the admissions to emergency for presumed Covid-19 cases, `nbre_hospit_corona`;
- the overall number of medical acts of SOS Médecins, `nbre_act_tot`;
- the number of medical acts of SOS Médecins of presumed Covid-19 cases, `nbre_act_corona`.

As can be noticed in Fig. 13, the number of admissions to emergency and the SOS Médecins acts, in Auvergne-Rhône-Alpes region, dropped around the middle of March, when the Covid-19 epidemic spread started in France and the lockdown was imposed by the French health authorities, on March 17th. Nevertheless, as can be remarked from Fig. 13, the number of presumed or overt Covid-19 cases started consistently growing at the same time in Auvergne-Rhône-Alpes region. Analogous data evolution can be observed in several French regions, where the Covid-19 pandemic spread relevantly.

Consider here the ith region, the index is omitted for notation simplicity. Suppose that the number of hospitalizations, that is the time series `incid_hosp` in Fig. 11 denoted here as $H(t)$, is proportional to the number of infected individuals at a previous instant:

$$H(t) = \mu f(t - \eta).$$

The time series $f(t)$ being known, parameters μ and η can be identified by simple least squares minimization, repeated for different

values of η. The same can be done for the other time series illustrated in Section 2, that is

- new hospitalizations, `incid_hosp`;
- new intensive cares patients, `incid_rea`;
- daily patient demises, `incid_dc`;
- recovered patients, `incid_rad`,
- positive tested cases, `pos`,
- daily demises certified to Inserm, `incid_inserm`,
- admissions to emergency for presumed Covid-19 cases, `nbre_pass_corona` (denoted `pass_corona` in the tables);
- new hospitalizations among the admissions to emergency for presumed Covid-19 cases, `nbre_hospit_corona` (denoted `hospit_corona` in the tables);
- medical acts of SOS Médecins of presumed Covid-19 cases, `nbre_act_corona` (denoted `act_corona` in the tables)

The results of the fitting for the three regions that experienced the most relevant pandemic spread, i.e. Île-de-France (11), Grand Est (44) and Auvergne-Rhône-Alpes (84) are shown in Figs. 14–18, Figs. 19–23, and Figs. 24–28, respectively, those concerning the other regions can be found in Fiacchini and Alamir (2020).

The optimal values of parameters η and μ for the three regions are given in Tables 2 and 3. It can be noticed that the curves of the excess of deaths are reproduced, scaled and shifted in time, in the measured time series. A coherent sequence of event could be inferred from the values of η. It seems, in fact, that the peak of SOS Médecins acts occurred first, followed three days later by the peaks at the emergency services; then the peaks at intensive care units happened followed by the peaks of hospitalizations. Afterward, the peak of the excess of deaths, i.e. of f_i, occurred, while the peaks of the deaths registered at hospitals and by

M. Fiacchini and M. Alamir
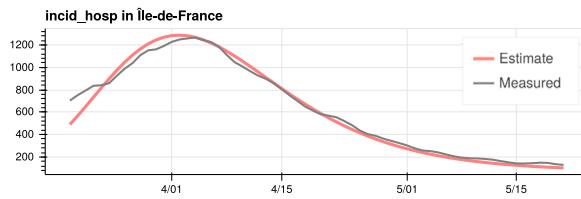
Fig. 14. Comparison of time series `incid_hosp` and `incid_rea` averaged over a 14-days-long window, with the fitted outputs, for Île-de-France (11) region.
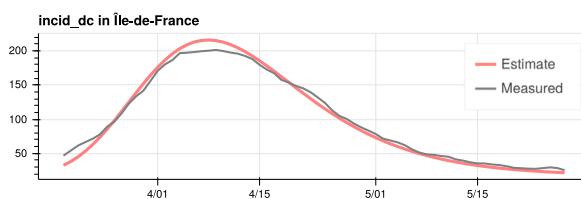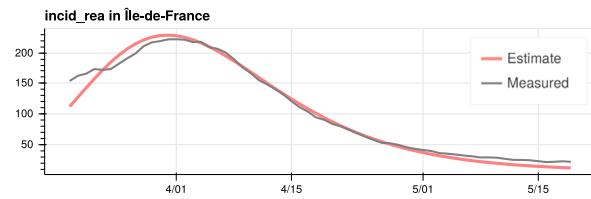

Fig. 15. Comparison of time series `incid_dc` and `incid_rad` averaged over a 14-days-long window, with the fitted outputs, for Île-de-France (11) region.
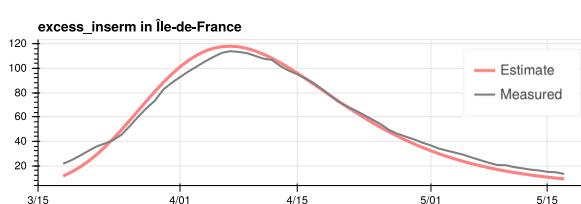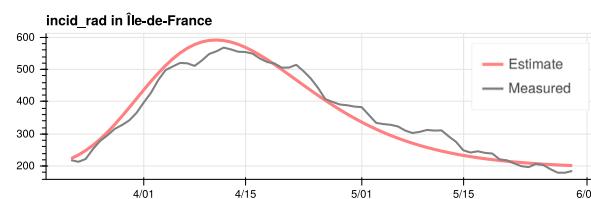

Fig. 16. Comparison of time series `incid_inserm` and `pos` averaged over a 14-days-long window, with the fitted outputs, for Île-de-France (11) region.
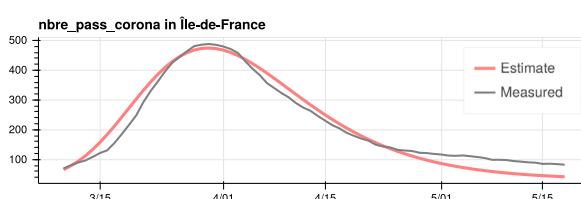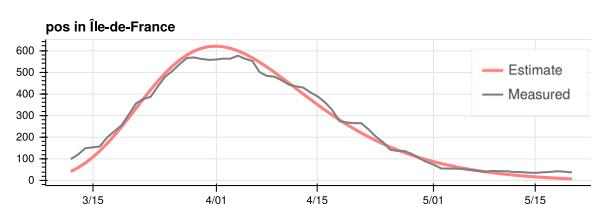

Fig. 17. Comparison of time series `nbre_pass_corona` and `nbre_hospit_corona` averaged over a 14-days-long window, with the fitted outputs, for Île-de-France (11) region.
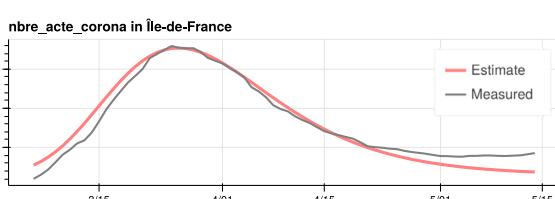

Fig. 18. Comparison of time series `nbre_acte_corona` averaged over a 14-days-long window, with the fitted output, for Île-de-France (11) region.

Table 2
Optimal parameters η and μ for Île-de-France (11), Grand Est (44) and Auvergne-Rhône-Alpes (84) regions.

	incid_hosp	incid_rea	incid_dc	incid_rad	incid_inserm
η_{11}	4	6	−2	−5	−1
μ_{11}	2.890469	0.530296	0.474127	0.951683	0.275693
η_{44}	3	5	−1	−5	0
μ_{44}	2.846007	0.524773	0.505064	0.966889	0.331267
η_{84}	5	6	−3	−3	−2
μ_{84}	3.430858	0.806820	0.399725	1.261361	0.343933

Inserm came, few days later. Finally, the curve of recovered patients arrived.

It is also interesting to note that, while the temporal sequence are substantially the same for Île-de-France (11) and Auvergne-Rhône-Alpes (84), in Grand Est (44) region, where the most important and early Covid-19 cluster revealed, the events appear to be compressed in time, the peaks at the emergency units, for SOS Médecins and for positive test occurred around 5 days before than the other two regions.

Moreover, from the comparison between the values of μ parameters, that are the proportionality coefficients between the evolution of excess of deaths and the considered time series, interesting information can be recovered, in particular on the degree of stress and saturation of the regional health systems. Considering for instance the acts of SOS Médecins, while there have been almost 2 acts for every demise of

Covid-19 in Auvergne-Rhône-Alpes , in Grand Est the number of deaths have been higher than the capability of the institution diagnosis, while in Île-de-France the number of demises seems to be more than the double of the SOS Médecins acts. Analogous conclusions can be drawn by other indicators: the activity of the emergency units (`pass_corona` and `hospit_corona`); the number of positive test; the hospitalization and intensive care accesses, in the three regions. It appear evident the degree of overwhelming saturation of the health systems of Grand Est and, even more dramatically, of Île-de-France region.

5. Conclusions

In this paper, the available data concerning the Covid-19 spread in the 13 different regions of mainland France have been considered to
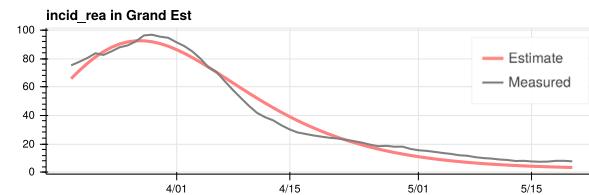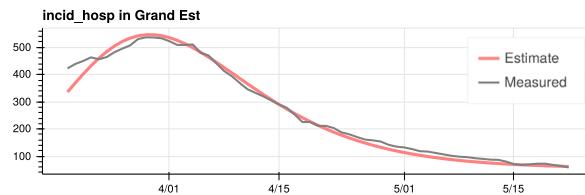
Fig. 19. Comparison of time series `incid_hosp` and `incid_rea` averaged over a 14-days-long window, with the fitted outputs, for Grand Est (44) region.
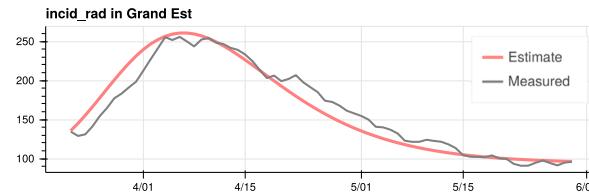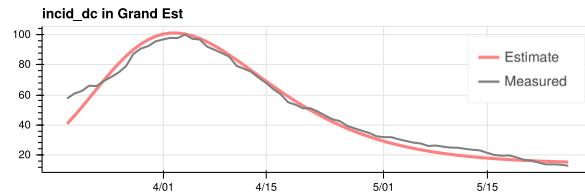

Fig. 20. Comparison of time series `incid_dc` and `incid_rad` averaged over a 14-days-long window, with the fitted outputs, for Grand Est (44) region.
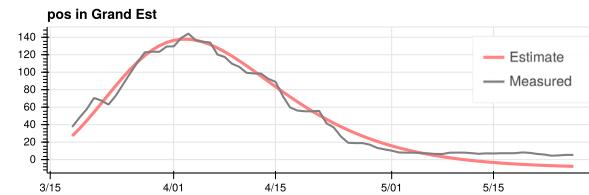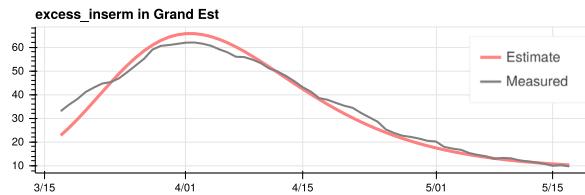

Fig. 21. Comparison of time series `incid_inserm` and `pos` averaged over a 14-days-long window, with the fitted outputs, for Grand Est (44) region.
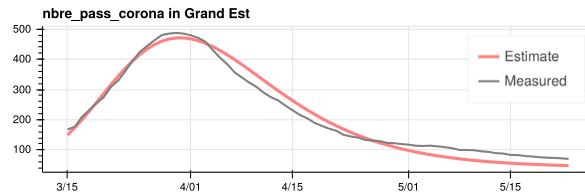

Fig. 22. Comparison of time series `nbre_pass_corona` and `nbre_hospit_corona` averaged over a 14-days-long window, with the fitted outputs, for Grand Est (44) region.

Table 3
Optimal parameters η and μ for Île-de-France (11), Grand Est (44) and Auvergne-Rhône-Alpes (84) regions.

|	pos	pass_corona	hospit_corona	act_corona
η_{11}	5	7	7	11
μ_{11}	1.496454	1.053246	0.470787	0.385416
η_{44}	−1	2	3	6
μ_{44}	0.853961	2.487286	1.116551	0.859851
η_{84}	5	6	7	10
μ_{84}	0.693771	5.060418	2.269344	1.831710
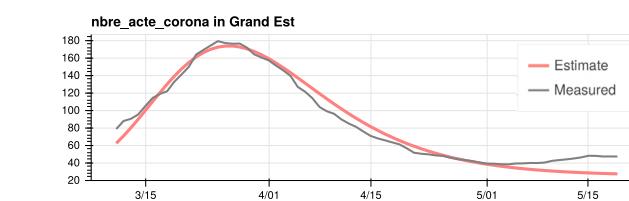
Fig. 23. Comparison of time series `nbre_acte_corona` averaged over a 14-days-long window, with the fitted output, for Grand Est (44) region.

build and identify a dynamical model. The aim was to fit the data with a model as simple as possible, to prevent and overcome the possible problem of nonidentifiability. An extremely simple dynamical system, with two states and only two common parameters for all the regions, permitted to reproduce the considered time series, that are the data concerning hospitalizations, intensive cares accesses, emergency and SOS Médecins activity, Covid-19 testing. These results seem to indicate that, with the available data only, more complex epidemiological models might incur in nonidentifiability and raise issues on their forecasting power. Exploiting additional data and further knowledge could be a solution to this problem.
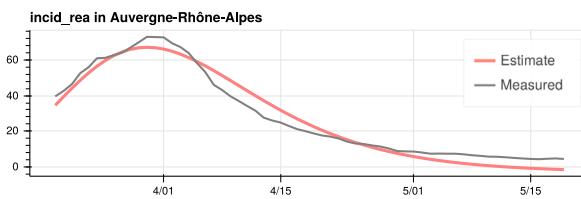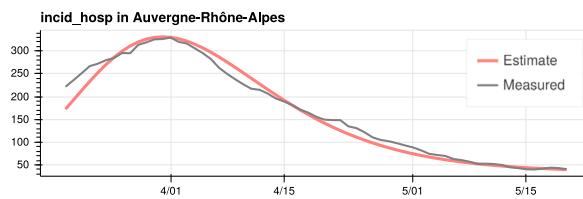
Fig. 24. Comparison of time series `incid_hosp` and `incid_rea` averaged over a 14-days-long window, with the fitted outputs, for Auvergne-Rhône-Alpes (84) region.
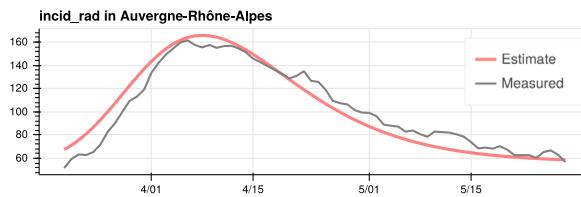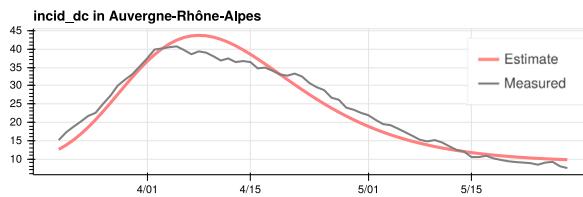

Fig. 25. Comparison of time series `incid_dc` and `incid_rad` averaged over a 14-days-long window, with the fitted outputs, for Auvergne-Rhône-Alpes (84) region.
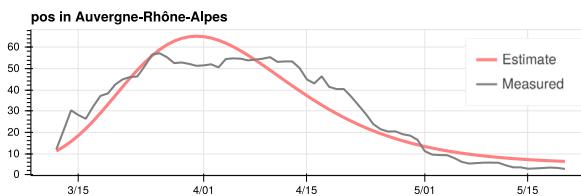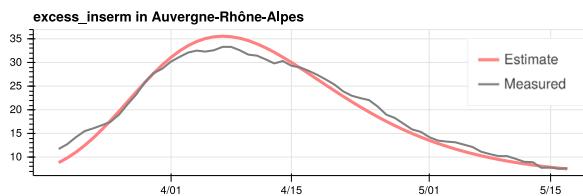

Fig. 26. Comparison of time series `incid_inserm` and `pos` averaged over a 14-days-long window, with the fitted outputs, for Auvergne-Rhône-Alpes (84) region.
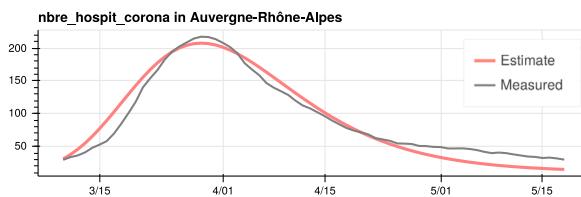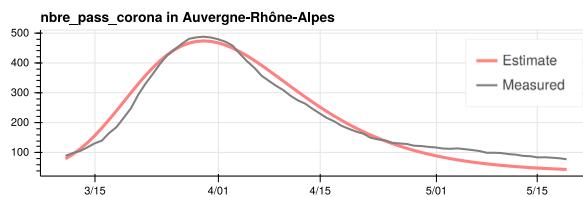

Fig. 27. Comparison of time series `nbre_pass_corona` and `nbre_hospit_corona` averaged over a 14-days-long window, with the fitted outputs, for Auvergne-Rhône-Alpes (84) region.
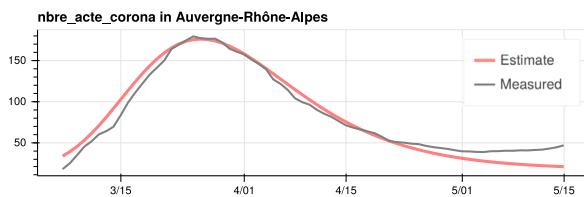

Fig. 28. Comparison of time series `nbre_acte_corona` averaged over a 14-days-long window, with the fitted output, for Auvergne-Rhône-Alpes (84) region.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

Alamir, M. https://github.com/mazen-alamir/torczon.

Alamo, T., Reina, D., & Millán, P. (2020). Data-driven methods to monitor, model, forecast and control Covid-19 pandemic: Leveraging data science, epidemiology and control theory. arXiv preprint arXiv:arXiv:2006.01731.

Brockmann, D., & Helbing, D. (2013). The hidden geometry of complex, network-driven contagion phenomena. *Science, 342*(6164), 1337–1342.

Burnham, K. P., Anderson, D. R., & Huyvaert, K. P. (2011). AIC model selection and multimodel inference in behavioral ecology: some background, observations, and comparisons. *Behavioral Ecology and Sociobiology, 65*(1), 23–35.

Cori, A., Ferguson, N. M., Fraser, C., & Cauchemez, S. (2013). A new framework and software to estimate time-varying reproduction numbers during epidemics. *American Journal of Epidemiology, 178*(9), 1505–1512.

Crokidakis, N. (2020). Data analysis and modeling of the evolution of COVID-19 in Brazil. arXiv preprint arXiv:arXiv:2003.12150.

El-Sayed, A. M., Scarborough, P., Seemann, L., & Galea, S. (2012). Social network analysis and agent-based modeling in social epidemiology. *Epidemiologic Perspectives & Innovations, 9*(1), 1.

Fang, Y., Nie, Y., & Penny, M. (2020). Transmission dynamics of the COVID-19 outbreak and effectiveness of government interventions: A data-driven analysis. *Journal of Medical Virology, 92*(6), 645–659.

Ferretti, L., Wymant, C., Kendall, M., Zhao, L., Nurtay, A., Abeler-Dörner, L., et al. (2020). Quantifying SARS-CoV-2 transmission suggests epidemic control with digital contact tracing. *Science, 368*(6491).

Fiacchini, M., & Alamir, M. (2020). The ockham's razor applied to COVID-19 model fitting french data. arXiv:2007.11379.

Gevertz, J., Greene, J., Tapia, C. H. S., & Sontag, E. D. (2020). A novel COVID-19 epidemiological model with explicit susceptible and asymptomatic isolation compartments reveals unexpected consequences of timing social distancing. medRxiv.

Giordano, G., Blanchini, F., Bruno, R., Colaneri, P., Di Filippo, A., Di Matteo, A., et al. (2020a). Modelling the COVID-19 epidemic and implementation of population-wide interventions in Italy. *Nature Medicine*, 1–6.

Giordano, G., Blanchini, F., Bruno, R., Colaneri, P., Di Filippo, A., Di Matteo, A., et al. (2020b). A SIDARTHE model of COVID-19 epidemic in Italy. arXiv preprint arXiv:arXiv:2003.09861.

Huang, N. E., Qiao, F., & Tung, K.-K. (2020). A data-driven model for predicting the course of COVID-19 epidemic with applications for China, Korea, Italy, Germany, Spain, UK and USA. medRxiv.

Kermack, W. O., & McKendrick, A. G. (1927). A contribution to the mathematical theory of epidemics. In *Containing Papers of a Mathematical and Physical Character Proceedings of The Royal Society of London. Series A, 115*(772), 700–721.

Kucharski, A. J., Russell, T. W., Diamond, C., Liu, Y., Edmunds, J., Funk, S., et al. (2020). Early dynamics of transmission and control of COVID-19: a mathematical modelling study. *The Lancet Infectious Diseases*.

McKinley, T., Cook, A. R., & Deardon, R. (2009). Inference in epidemic models without likelihoods. *International Journal of Biostatistics, 5*(1).

Morato, M. M., Bastos, S. B., Cajueiro, D. O., & Normey-Rico, J. E. (2020). An optimal predictive control strategy for COVID-19 (SARS-CoV-2) social distancing policies in Brazil. arXiv:2005.10797.

Nishiura, H., Klinkenberg, D., Roberts, M., & Heesterbeek, J. A. (2009). Early epidemiological assessment of the virulence of emerging infectious diseases: a case study of an influenza pandemic. *PLoS One, 4*(8), Article e6852.

Portet, S. (2020). A primer on model selection using the Akaike information criterion. *Infectious Disease Modelling, 5*, 111–128.

Postnikov, E. B. (2020). Estimation of COVID-19 dynamics "on a back-of-envelope": Does the simplest SIR model provide quantitative parameters and predictions? *Chaos, Solitons & Fractals, 135*, Article 109841.

Qian, G., & Mahdi, A. (2020). Sensitivity analysis methods in the biomedical sciences. *Mathematical Biosciences, 323*, Article 108306.

Roda, W. C., Varughese, M. B., Han, D., & Li, M. Y. (2020). Why is it difficult to accurately predict the COVID-19 epidemic? *Infectious Disease Modelling.*

Wu, J. T., Leung, K., & Leung, G. M. (2020). Nowcasting and forecasting the potential domestic and international spread of the 2019-nCoV outbreak originating in Wuhan, China: a modelling study. *The Lancet, 395*(10225), 689–697.